

特開平8-146987

(43) 公開日 平成8年(1996)6月7日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 1 3 Z			
	5 7 1 G			
G 0 6 F 3/16	3 2 0 F	9172-5E		

審査請求 未請求 請求項の数 8 O L (全 13 頁)

(21) 出願番号	特願平6-283260	(71) 出願人	000001007 キヤノン株式会社 東京都大田区下丸子3丁目30番2号
(22) 出願日	平成6年(1994)11月17日	(72) 発明者	山本 寛樹 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
		(72) 発明者	小森 康弘 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
		(72) 発明者	山田 雅章 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
		(74) 代理人	弁理士 大塚 康徳 (外1名)

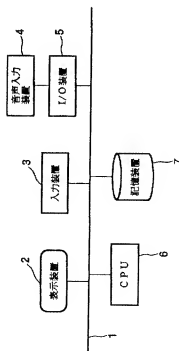
最終頁に続く

(54) 【発明の名称】 音声入力装置及びその制御方法

(57) 【要約】

【目的】 有効な音声情報を会処理に渡すために、音声入力する状況に応じて、音声検出する方法を変更することを実現する。

【構成】 音声入力装置4から入力された音声部分の始点と終点を特定するモードとして、自動的にそれらを特定するモードと、始点位置のみをマニュアルで指定し、終点を自動にするモード、更には、始点及び終点の両方をマニュアルで指定するモードの3つあって、それらのいずれかを入力装置3より選択する。



【特許請求の範囲】

【請求項1】 音声入力手段から入力された音声情報を下位の処理に渡す音声入力装置であって、

前記音声入力手段から入力される有意な音声情報の期間を特定するための複数のモードと、

該複数のモードの中から1つを選択する選択手段と、

を備えることを特徴とする音声入力装置。

【請求項2】 前記モードには、音声の始点及び終点を自動識別する第1のモード、始点位置をマニュアル指示する第2のモード、始点及び終点位置をマニュアル指示する第3のモードが含まれることを特徴とする請求項第1項に記載の音声入力装置。

【請求項3】 前記選択手段は、初期段階では前記第1のモードを選択し、所定の変更指示があるとき前記第2のモード、第3のモードの順に選択することを特徴とする請求項第2項に記載の音声入力装置。

【請求項4】 前記選択手段は、マニュアルにより選択することを特徴とする請求項第1項に記載の音声入力装置。

【請求項5】 音声入力手段から入力された音声情報を下位の処理に渡す音声入力装置の制御方法であって、前記音声入力手段から入力される有意な音声情報の期間を特定するための複数のモードと、

該複数のモードの中から1つを選択する選択工程と、

を備えることを特徴とする音声入力装置の制御方法。

【請求項6】 前記モードには、音声の始点及び終点を自動識別する第1のモード、始点位置をマニュアル指示する第2のモード、始点及び終点位置をマニュアル指示する第3のモードが含まれることを特徴とする請求項第5項に記載の音声入力装置の制御方法。

【請求項7】 前記選択工程は、初期段階では前記第1のモードを選択し、所定の変更指示があるとき前記第2のモード、第3のモードの順に選択することを特徴とする請求項第6項に記載の音声入力装置の制御方法。

【請求項8】 前記選択手段は、マニュアルにより選択することを特徴とする請求項第5項に記載の音声入力装置の制御方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は音声入力装置及びその制御方法、詳しくは入力された情報中の音声区間を特定し、下位の処理に渡す音声入力装置及びその制御方法に関するものである。

【0002】

【従来の技術】 一般に、音声入力インタフェースは、音声認識等の他のアプリケーションにユーザの発声内容を渡す処理を行う。そのため、入力される音声データから、ユーザが発声した音声区間を検出することが必要になってくる。

【0003】 音声区間を検出する方法は様々あり、例え

ば次に示す方法がある。

検出方法1：ユーザがマウスやキー等の入力装置を用いて、発声の開始点と終了点を決定する。

検出方法2：発声の開始点はユーザが決め、終了点は計算機が判断する。

検出方法3：発声の開始点は計算機が判断し、終了点はユーザが決める。

検出方法4：発声の開始点、終了点ともに計算機が判断する。

【0004】 また、上記した各音声区間検出方法に対応する、音声入力の際にユーザに要求される操作の例を次に示す。ただし、ここでは入力装置としてマウスを用いた場合である。

操作方法1：発声する前にマウスボタンを押し、発声中は押し続け、発声終了後マウスボタンを離す。

操作方法2：発声する前に一度だけマウスボタンを押下する。

操作方法3：発声終了後に一度だけマウスボタンを押下する。

【0005】 操作方法4：ユーザの操作は必要ない。

【0006】 従来の音声入力インタフェースでは上記した検出方法のいずれか1つの手法のみを備えていた。

【0007】

【発明が解決しようとする課題】 前記した各検出方法及び操作方法にはそれぞれ長所と短所がある。

【0008】 例えば、検出方法1では音声区間の始端・終端をユーザが与えるため、音声区間の切り出し間違いが少ない。しかし、前記操作方法1に示した通り操作が面倒である。

【0009】 また、検出方法2、3では、検出方法1と比較し、ユーザの負担がやや軽くなるが、その一方で、計算機が雑音を音声と誤って検出し、検出された音声区間に雑音が含まれることがある。

【0010】 検出方法4ではユーザに全く操作を要求しないため使用感は向上するが、やはり音声入力環境の変化や雑音などにより、発声した音声を検出しなかったり、雑音を音声として検出してしまったりすることがある。

【0011】 このように従来の音声入力インタフェースでは、検出方法のいずれか1種の方法のみを用いているため、検出方法のもつ短所がそのまま音声入力インタフェースの課題の一つになっている。

【0012】

【課題を解決するための手段】 及び

【作用】 本発明はかかる問題点に鑑みなされたものであり、音声入力する状況に応じて検出方法を変更することを可能にし、有効な音声情報を下位処理に渡す音声入力装置及びその制御方法を提供しようとするものである。

【0013】 この課題を解決するため、例えば本発明の音声入力装置は以下の構成を備える。すなわち、音声入

3

力手段から入力された音声情報を下位の処理に渡す音声入力装置であって、前記音声入力手段から入力される有意な音声情報の期間を特定するための複数のモードと、該複数のモードの中から1つを選択する選択手段とを備える。

【0013】また、本発明に係る好適な実施態様に従えば、前記モードには、音声の始点及び終点を自動識別する第1のモード、始点位置をマニュアル指示する第2のモード、始点及び終点位置をマニュアル指示する第3のモードが含まれることが望ましい。これによって、ユーザの置かれている状況に応じたモードを網羅することが可能になる。

【0014】また、前記選択手段は、初期段階では前記第1のモードを選択し、所定の変更指示があると前記第2のモード、第3のモードの順に選択することが望ましい。この結果、通常の使用環境においては、ユーザに全く負担のかけられないモードが選択される。また、使用環境が変化した場合には、環境の変化に応じてモードが適宜変更される。

【0015】また、選択手段は、マニュアルにより選択することによっても良い。これによれば、直ちに、ユーザの置かれている環境に適応させることが可能になる。

【0016】

【実施例】以下、添付図面に従って本発明に係る実施例を詳細に説明する。図1は本発明に係る情報処理装置である計算機の第1実施例の概略構成を表すブロック図である。

【0017】図中、1はシステムバスであり、このシステムバス1には、CRTディスプレイ等の表示装置2、キーボードやマウス等の入力装置3、マイクロフォン等の音声入力装置4、前記音声入力装置4から供給される音声信号を計算機で処理できるデータに変換するI/O装置5、システム全体の動作を制御するCPU6が接続されている。尚、このCPU6内には、後述するフローチャートに従った処理を実現するためのプログラムを格納したROM、及びワークエリアとして使用するRAMで構成される主メモリを内蔵しているものとする。

【0018】さて、以上のような構成におけるシステムにおいて、例えば、実施例では、入力装置3としてマウスを用いていることにする。また、3種類の音声入力モードがある場合を考え、それぞれのモードでユーザに要求される操作とそれぞれのモードに対応する音声区間の抽出方法は次のようなものとする。

【0019】音声入力モード

モード1：ユーザの操作を必要としない。

モード2：ユーザは発声する前にマウスを一度だけ押下する。

モード3：ユーザは発声する前にマウスを押し、発声中マウスを押し続け、発声終了後にマウスから手を放す。

4

【0020】従って、各モードにおける音声区間検出方法は次のようになる。

モード1：音声区間の始端、終端の検出を計算機が行なう。

モード2：音声区間の始端をユーザが決定し、終端は計算機が検出する。

モード3：音声区間の始端、終端をユーザが決定する。

【0021】システムの動作を図2のフローチャートに従って説明する。

10 【0022】まず、ステップS1において、音声入力モードを選択する。図3に、この音声入力モード選択処理の一例を示す。

【0023】まず、ステップS11で、選択可能音声入力モードを表示し、ステップS12でユーザによる選択を持つ。ユーザはマウス（入力装置3）を用いて所望の入力モードを選択する。ユーザが音声入力モードの1つを選択すると、ステップS13で選択入力モードを確認させるためにユーザに知らせると共に、その選択内容を所定の記憶領域（実施例ではCPU6内の主記憶メモリの所定アドレス）に記憶保持させる。

20 【0024】上記の如く、音声入力モードが選択されると、処理は図2のステップS2に進み、時刻tを0に設定し、また、音声区間の始端・終端の検出状況を示す変数Stを、始端・終端ともに検出されていないことを示すNOTYETに設定する。尚、時刻tを“0”にセットするのは、不図示のタイマをリセットするものであり、変数StはCPU6内の主メモリ内の所定アドレスに確保されているものである。この初期設定処理後、ステップS3の音声取り込み処理、ステップS5の音声データ分析処理に移る。

30 【0025】音声取り込み処理（ステップS3）では、ユーザの発した音声マイクロフォンなどの音声入力装置4とI/O装置5を用いて、計算機の処理できるデータに変換し、計算機に取り込む。次に、音声データ保持処理（ステップS4）では、前記ステップS3で取り込まれた音声データを記憶装置7に保持する。音声データを保持後、再び音声取り込み処理に戻り同様の動作を繰り返し、音声取り込みと音声データの保持を続ける。

40 【0026】前記ステップS3、S4と並行して、ステップS5の音声データ分析処理が行われ、前記ステップS4で記憶装置7に保持された音声データから、予め定められた時間幅 Δt の分量だけデータを読み込み音声データの分析を行なう。音声データの分析処理の一例を図4のフローチャートに示す。

【0027】図4において、ステップS51で時刻tから時刻t+ Δt 間の音声データを記憶装置7から読み込む。続くステップS52では、読み込まれた音声データの二乗平均値を計算して、それをP（t）として求める。

【0028】図2に戻り、続くステップS6ではステッ

5

ブS1で設定した音声入力モードに対応する音声区間検出を行なう。ここで、ステップS1でユーザが音声入力モードとしてモード1を選んだ場合、音声区間検出の一例を図5のフローチャートに、モード2を選んだ場合の音声区間検出の一例を図6のフローチャートに、モード3を選んだ場合の音声区間検出の一例を図7のフローチャートに示す。

【0029】まず、モード1が選択された場合の音声区間検出処理を説明する。

【0030】図5において、音声未入力判断ステップS61で音声区間検出の状況を示すStがNOTYETの場合（音声区間の始端が決定していない場合）、ステップS621に進んで、音声区間の始端であるかどうかを判定する。また、NOTYET以外の場合は、ステップS64に進む。

【0031】ステップS621では、先に求めた二乗平均値計算結果である、時刻tから時刻t+Δtにおける音声データの二乗平均値P(t)と、予め定められた音声の始端を判定するための閾値Tpsと比較し、P(t)が閾値Tpsよりも大きい場合はステップS63に進んで、始端を決定する。ここでは、Stを音声区間の始端が検出されたこと（すなわち音声が入力されたこと）を示す“IN”に変更し、時刻tを音声区間の始端時刻Stimeとし、音声区間検出の処理を終る。また、ステップS621でP(t)が閾値Tpsを越えないと判断した場合は、音声区間検出の処理を終了する。

【0032】一方、ステップS64に処理が進んで、音声入力期間中であると判断した場合には、StがINである場合（音声区間の始端が決定している場合）は音声区間の終端を判定する終端検出ステップS651に進み、StがIN以外である場合は音声区間検出の処理を終える。ステップS651の終端検出処理では、前記ステップS52で算出したP(t)と予め定められた音声の終端を判定するための閾値Tpeと比較する。P(t)がTpeよりも小さい場合はステップS66に進み、P(t)が閾値Tpe以上の場合は、音声区間検出の処理を終了する。

【0033】ステップS66では、Stを音声区間の終端が検出されたこと（すなわち音声入力終了したこと）を示すENDに変更し、時刻tを音声区間の終端時刻ETimeとして、音声区間検出の処理を終了する。

【0034】次に、図2におけるステップS1でモード2が選択された場合の音声区間検出処理を図6のフローチャートに従って説明する。

【0035】モード1の場合と異なる点は、音声未入力判断ステップS61で音声区間検出の状況を示すStがNOTYETの場合の処理である。その他の処理で同じ部分に関しては図5と同一符号を付した。

【0036】モード2では、ステップS61における音声未入力判断処理でStがNOTYETの場合、ユーザ

6

が音声区間開始を知らせるマウス押下を行なったか否かを判断するステップS622に進み、始端入力検出処理を行う。このステップでユーザが音声区間開始の合図であるマウス押下を行なったか否かを調べ、ユーザがマウスを押下していた場合には、ステップS63に進んで始端決定を行い、押下していない場合は、音声区間検出の処理を終る。その他のステップについては、モード1の場合と同じ処理を行なう。

【0037】次に、図2におけるステップS1でモード3が選択された場合の音声区間検出処理を図7のフローチャートに従って説明する。

【0038】モード2の場合と異なる点は、ステップS64における音声入力期間中判断処理で音声区間検出の状況を示すStがINの場合の処理である。

【0039】ステップS64において、StがINである場合は音声区間の終端を判定するステップS652に進み、StがIN以外である場合は音声区間検出の処理を終る。

【0040】ステップS652では、マウスボタンが（押下されていた状態から）開放されたか否かを調べ、マウスが開放された場合には、ステップS66に進んで終端を決定し、マウスが開放されていない（ユーザがマウスを押し続けている）場合は音声区間検出処理を終了する。その他のステップについては、モード2の場合と同じ処理を行なう。

【0041】再び、図2に戻って、上記の如く、音声区間検出処理が終了した後、ステップS7に進んで、音声入力終了判断を行う。ここでは、StがENDであればステップS9に進んで、音声区間を表示し、StがEND以外であればステップS8に進んで、分析時刻更新を行い、時刻tをΔtだけ増加する。そして、ステップS5に戻って音声データ分析を続行する。

【0042】ステップS9における音声区間表示処理では、上記ステップS66で検出した音声区間（時刻Stimeから時刻ETimeまで）を表示装置2に表示したのち、音声データを抽出する。

【0043】以上のごとく説明した実施例を、ユーザが突発的な雑音の発生する環境下で音声入力する場合に適用すると次のようになる。

【0044】まず、プログラム開始と同時にステップS1の音声入力モード選択処理で、表示装置2に選択可能な入力モードを表示し（ステップS11）、ユーザによる選択を待つ（ステップS12）。ここで、例えばユーザがモード1を選択したとすると、続くステップS13では、選択されたモードがモード1であることを、表示装置2上に表示するなどしてユーザに知らせる。音声入力モードの選択を表示装置2上に実現した一例を図8に示す。

【0045】同図では、各音声入力モードに呼称を付けている。それぞれの呼称は、モード1がKeep Pressing

SpeechInput, モード2がOne Click SpeechInput, モード3がHand Free SpeechInputである。この図では、モード3に対応するHand Free SpeechInput が選択されていることを示している。また、ステップS13でユーザに選択されたモードを伝えるため表示装置2上に表示した例を図9に示す。図示の如く、画面上部に選択モードを明示することで、現在のモードをユーザに知らせている。

【0046】続いてステップS2の初期設定処理に移行し、時刻 t 、音声検出状況 S_t の初期化を行なう。続く、ステップS3の音声声取り込み、ステップS4の音声データ保持により、音声取り込みと記憶装置2への音声データ保持が開始され、ステップS3、S4は繰り返され、同時にステップS5における音声分析処理では、記憶装置2に保持された音声データを定められた時間幅 Δt で分析し、ステップS6の音声区間検出処理へ移行する。

【0047】今、音声入力モードとしてモード1が選択されている例を説明しているので、図5にフローチャートを示した音声区間検出を行なうことになる。ステップS6の音声区間検出処理を終えると、音声区間検出状況 S_t を調べ、音声区間の終端が検出されるまで（音声区間が終了とみなされるまで）、時間幅 Δt ごとに音声データ分析処理（ステップS5）と音声区間検出処理（ステップS6）を繰り返して行なう。

【0048】説明のため、ユーザが音声入力モードを選択した後、ユーザが発生する前に、突発的な雑音（瞬時的に二乗平均値が閾値 Tps を上回る大きな音）が時刻 t_a から時刻 t_{a2} （ $t_{a2} > t_a$ ）にかけて発生したとする。この場合、時刻 t_a の音声データを処理する音声区間検出処理（ステップS6）では、突発的な雑音の二乗平均値が閾値 Tps より大きいため、雑音の始端が音声区間の始端として検出され、ステップS63の始端検出処理に移る。前記ステップS63では、 S_t を1Nに変更し、音声区間の始端の時刻 $Stime$ を t_n とする。

【0049】その後、突発的な雑音が弱まる時刻 t_{a2} における音声区間検出処理（ステップS6）の終端検出処理（ステップS651）で、雑音が弱まっているために、 $P(t)$ が閾値 Tpe よりも小さくなるため、終端決定を行う（ステップS66）。つまり、 s_t をENDに変更し、 t_{a2} を音声区間の終端の時刻 $EndTime$ とする。

【0050】計算機は突発的な雑音を音声と誤認して、雑音の発生していた時間（時刻 $t_a \sim t_{a2}$ 間）を音声区間として検出して、ステップS9で音声区間を表示する。ユーザはこの表示により、誤った音声区間を検出したことを確認できる。音声入力モードをモード1に設定して音声入力を行ない、突発的な雑音が音声区間を誤検出した一例を図10に示す。図10は、音声入力インタフェースの波形表示部の一例であり、中央の2層の白色

の表示部のうち、上の表示部が音声の原波形を示し、下の表示部は二乗平均値を表示している。また、上の表示部において、領域Aが検出された音声区間である。

【0051】2回目の実行では（前回の実行で、突発的な雑音が発生する環境下では、モード1の音声区間の始端検出が誤検出することが分かっている）ので、ステップS1の音声入力モード選択処理でユーザはモード2等を選択する。

【0052】以降の処理はモード1を選んだ際と同じで、異なる点は、音声区間検出で行われる処理である。前記ステップS12で音声入力モードはモード2を選択しているため、モード2に対応する図6にフローチャートを示した音声区間検出を行なう。

【0053】前回同様突発的な雑音が時刻 t_a から時刻 t_{a2} （ $t_{a2} > t_a$ ）にかけて再び発生したとする。前回の場合と異なり、時刻 t_a の音声データを処理する音声区間検出処理（ステップS6）では、ステップS62の始端入力検出より、どのような大きな雑音（あるいは音声）であっても、ユーザがマウスを押下するまでは音声区間の始端を検出したことにはならぬ。すなわち、ユーザがマウスを押下した時点まで音声区間の始端とするため、モード1を用いた前回のように、雑音に反応して音声区間の始端と誤判断することはなくなる。音声区間の始端を検出した以降の処理は前回と同様である。

【0054】突発的な雑音が発生している環境下で、音声入力モード2に設定して、ユーザがマウスを押下した後に発生し、正しく音声区間検出を行なわれた一例を図11に示す。図11において、ユーザが発生する前に突発的な雑音が入力されているが、ユーザがマウスを押下していなかったため、音声区間として検出されなかった。

【0055】以上の如く、突発的な雑音が入力された場合は音声入力モードをモード2に選択することで対応できた。

【0056】しかし、定めた閾値を定常的に上回る雑音が発生する環境ではモード1、モード2の音声区間検出では、前記終端検出処理（ステップS651）で常に $P(t) > Tpe$ となり、音声区間の終端検出が検出されない。このような環境下で音声入力する場合は、図7のフローチャートに示した音声区間検出を行なう方法モード3に切り替えることで対応できる。

【0057】モード3では、音声区間の始端、終端ともにユーザからマウスで指示されるため、雑音が発生しても正しく音声区間を検出できる。図12は音声入力モードをモード1に設定した場合に、定常的な雑音により音声区間が正しく検出されなかった一例である。これに対し、音声入力モードをモード3に設定して定常的な雑音環境下で音声区間を検出した一例を図13に示す。

【0058】以上の如く、本実施例によれば、音声入力区間を決定するモードをユーザが選択できるので、その

ユーザの置かれた状況に適応して検出方法を変更し、正しく音声区間を検出することが可能になる。

【0059】因みに、ユーザの操作が簡便なのは、モード1であり、次いで、モード2、モード3と続くが、ユーザが自身の置かれた状態、或いは環境に応じたモードを選択することができるので、操作性と正しく音声区間を検出する検出方法の両方を最善のものとすることが可能になる。

【0060】【第2の実施例の説明】図14のフローチャートに従って第2の実施例における動作処理内容を説明する。尚装置構成は図1と同様であるものとする。

【0061】さて、本第2の実施例では、第1実施例と同様の3種の音声入力モードを持つ音声入力インタフェースについて、計算機が自動的に入力モードを設定して、音声入力を行ない、設定された入力モードに対応する音声区間検出の結果をユーザに示す。ユーザは示された結果から、音声区間検出が正しく行なわれているか否かを判断する。正しく区間検出が行なわれていない場合は計算機が音声入力モードを変更し、正しく行なわれている場合は入力モードを変更しない。

【0062】詳細を図14と図15のフローチャートを用いて説明する。

【0063】第1実施例と異なる処理を行なうのは、図14における、ステップS1' における音声入力モード選択処理と、ステップS9' の音声区間表示処理後に行なうステップS10のキャンセル判断処理である。その他の各ステップは図2のと同様である。

【0064】ステップS1' の音声入力モード選択処理の例を図15のフローチャートを用いて説明する。

【0065】図15において、まず、ステップS101で起動直後判断処理では、プログラム起動直後であるか否かを判断する。起動直後である場合にはステップS102に進んで、音声入力モードとしてモード1を選択する。そして、選択入力モード伝達ステップS13' (図3のステップS13と同様)に進む。

【0066】また、起動直後ではない場合はステップS103に進み、現在のモードがモード1であるかどうかを判断する。現在の音声入力モードの設定がモード1の場合は、ステップS104に進んで、音声入力モードとしてモード2に設定し、ステップS13' の選択入力モード伝達処理を行う。

【0067】更に、現在のモードがモード1でもないとして判断した場合には、処理はステップS103からステップS105に進み、現在の音声入力モードの設定がモード2であるかどうかを判断する。モード2であると判断した場合には、ステップS106に進み、音声入力モードをモード3に設定して、ステップS13' の入力モード伝達処理をおこなう。

【0068】また、設定されているモードがモード2でもない、すなわち、モード3である場合には、ステップ

S11' に進み、選択可能な入力モードを表示装置2上に表示し、ステップS12' でユーザの選択を待つ。ユーザが音声入力モードを設定した後に前記ステップS13' に移る。

【0069】以上のようにして、モードが決定されると、図14に戻り、第1実施例と同様の処理を行なう。そして、ステップS9' で区間検出結果を表示し、ステップS10でキャンセル判断を行う。ユーザは発声した音声に対して正しく音声区間検出されている場合はキャンセルしない。誤って雑音等を音声区間として誤検出している場合はキャンセルする。ステップS10のキャンセル判断処理では、ユーザがキャンセルしたか否かを判断し、キャンセルした場合には前記ステップS1' に移り、キャンセルしない場合にはステップS5に進む。キャンセルは入力装置3を用いて行ない、その方法は予め決めておく。例えば、キャンセルの方法としてはマウスをダブルクリックする、ないしはキーボード上の特定のキーを押下する等である。

【0070】以上に示した実施例を実際にユーザが使用した場合について説明する。

【0071】まず、プログラム起動と同時にステップS102で音声入力モードがモード1 (すなわち、一番操作が簡便なモード) に設定され、ステップS13' でモード1が選択されたことをユーザに知らせる。ステップS2' で初期設定後、ステップS3'、S4' で音声取り込み及び音声データ保持が行なわれ、ステップS5' である時刻tから時刻t+Δtの音声データを分析し、ステップS6' で音声区間検出処理が行なわれる。ステップS7' で音声検出状況StがENDにならない間は、次の時刻の音声データを分析し、音声区間検出および音声入力終了を判断するステップS5' ~S8' を繰り返す。ステップS7' で音声入力終了と判断した場合は音声区間表示を行なうステップS9' に移り、検出された音声区間を表示する。続いて、ステップS10のキャンセル判断処理に移る。

【0072】ここで音声区間が正しく検出されたとすると、ステップS10において、ユーザがキャンセルしないのでステップS2' に進み、同じ音声入力モード(モード1)で音声入力を行なう。ここで、音声入力の際に、突発的な雑音が、ユーザの発声前に生じ、音声区間を誤検出したとする。ユーザはステップS9' により、雑音を誤って検出したことを認識できるので、続くステップS10でキャンセルする。ステップS10でキャンセルされた場合は、ステップS11に進んで、音声入力モード選択処理を行い、図15のフローチャートに示した処理が行なわれる。この場合、現在の音声入力モードはモード1であるから、ステップS101、S103を経た後、ステップS104でモード2が設定される。また、モード2に変更されたことがステップS13' により、ユーザに知らされる。続いて、モード2の音声入力

モードで音声入力が行なわれる。

【0073】尚、例えば、モード2の音声入力モードを用いても、定期的に大きな雑音が入り、音声区間の終端が正しく検出されなかった場合は、ユーザがステップS10でキャンセルすることにより、ステップS11'の音声入力モード選択処理に戻り、ステップS101、S103、S105を経て、ステップS106でモード3に変更される。モード3で入力している場合は、ユーザが音声区間を設定しているので雑音などによる誤検出はなくなる。

【0074】モード3で音声入力中に音声入力環境が変化し、雑音が小さくなり、操作の容易な入力モードに変える場合は、ユーザがキャンセルすることによってユーザが任意の入力モードを設定できるようにする。この場合、ステップS10から音声入力モード選択処理（ステップS11'）に移り、ステップS101、S103、S105を経て、ステップS11'で選択可能な音声入力モードが表示される。ユーザは続くユーザ選択処理（ステップS12'）で任意の入力モードを選択し音声入力を続けることができる。

【0075】以上では、音声入力モードを自動選択する一実施例を説明した。上記説明の通り、プログラム起動直後はユーザの操作を必要としない音声入力モードで音声入力を行ない、ユーザがキャンセルする毎に、段階的に区間検出をユーザに依存した入力モードに変更していく方法である。

【0076】なお、本発明は、図示の実施例に限定されず、種々の変形が可能である。例えばその変形例には次のようなものがある。

(1) 上記実施例では、入力装置2としてマウスを用いたが、これに限定されず、キーボード上のキーやライトペン、タッチパネル等を用いても良い。

(2) 上記実施例では、計算機で音声区間検出をする際に二乗平均値をパラメータとして用いていたが、これに限定されず、零交差回数を窓をかけたパワー等を用いても良く、またこれらのパラメータを複数用いても良い。

(3) 上記実施例では、3種の音声区間検出方法を用いたが、これに限定されず、例えば、

(a) 計算機、ユーザともに音声区間の始端（終端）とした場合を音声区間の始端（終端）とする。

【0077】(b) 計算機が始端（終端）と判断した時刻付近でユーザが入力装置2に始端（終端）を知らせる操作を行なった場合を始端（終端）とする。等を用いても良い。

(4) 上記実施例では、ユーザに選択された音声入力モードを伝達する方法として、画面上に選択されたモード名を表示したが、これに限らず、合成音を用いて選択されたモードをユーザに知らせても良い。

【0078】以上説明したように、本実施例によれば、複数の音声入力モードの中から音声入力モードを設定す

る音声入力モード設定手段と、ユーザの発生内容を取り込み保持する音声取り込み保持手段と、複数の音声区間検出方法から該音声入力モード設定手段により設定された音声入力モードに対応する音声区間検出を行なう音声区間検出手段とを備えたことにより、音声入力環境の変化や雑音等に応じて、複数の音声入力モードから音声入力モードを選択でき、操作性、使用感が著しく向上する。

【0079】特に、第2の実施例によれば、初期段階では操作性が簡便なモードが選択されるので、比較的雑音の少ない場所等の環境では一番ユーザに負担がかからないモードにすることが可能になる。

【0080】また、上記実施例では、1つの独立した装置に適応した例を説明したが、上記の説明から容易に推察されるごとく、本発明は、複数の機器から構成されるシステムに適用しても構わない。また、本発明はシステム或は装置にプログラムを供給することによって達成される場合にも適用できることは言うまでもない。

【0081】

【発明の効果】以上説明したように本発明によれば、音声入力する状況に応じて音声を検出方法を変更し、有効な音声情報を下位処理に渡すことが可能になる。

【0082】

【図面の簡単な説明】

【図1】本発明に係る情報機器の第1実施例のブロック図である。

【図2】第1の実施例のメイン処理を示すフローチャートである。

【図3】第1の実施例における音声入力モード選択の処理を示すフローチャートである。

【図4】第1の実施例における音声データ分析処理を示すフローチャートである。

【図5】第1の実施例における音声区間検出のモード1の処理を示すフローチャートである。

【図6】第1の実施例における音声区間検出のモード2の処理を示すフローチャートである。

【図7】第1の実施例における音声区間検出のモード3の処理を示すフローチャートである。

【図8】第1の実施例において音声入力モードを選択する際に選択可能な音声入力モードを表示した例である。

【図9】第1の実施例において、選択された音声入力モードをユーザに伝えるために、画面上に表示した例である。

【図10】第1の実施例において、モード1の音声区間検出処理が突発的な雑音の発生により、音声区間を誤検出した例である。

【図11】第1の実施例において、モード2の音声区間検出処理が、音声区間を正しく検出した例である。

【図12】第1の実施例において、モード1の音声区間検出処理が定期的な雑音の発生により、音声区間を検出

できなかった例である。

【図13】第1の実施例において、モード3の音声区間検出処理が突発的な雑音の発生により、音声区間を正しく検出した例である。

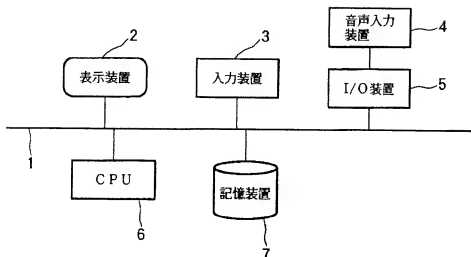
【図14】第2の実施例のメイン処理を示すフローチャートである。

【図15】第2の実施例における音声入力モード選択の処理を示すフローチャートである。

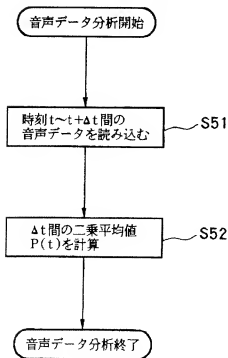
【符号の説明】

- 1 システムバス
- 2 表示装置
- 3 入力装置
- 4 音声入力装置
- 5 I/O装置
- 6 CPU
- 7 記憶装置

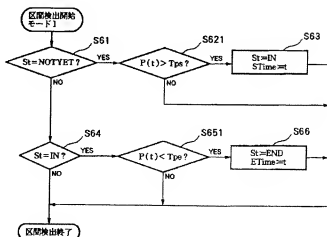
【図1】



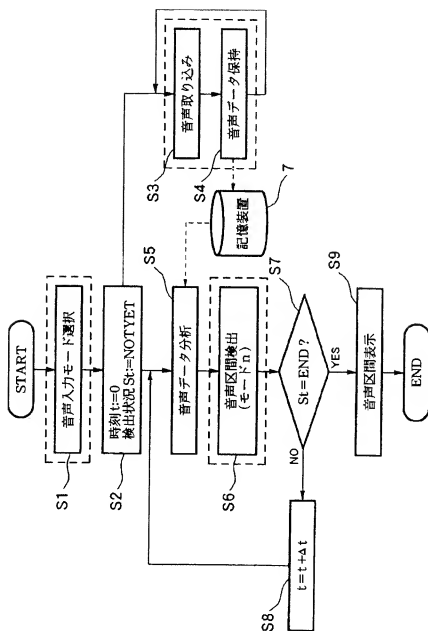
【図4】



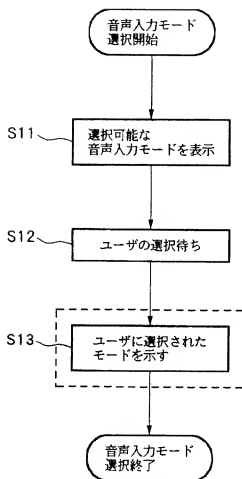
【図5】



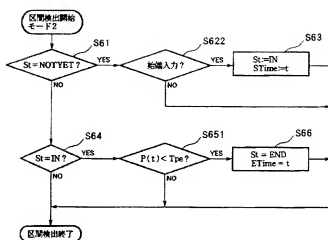
【図2】



【図3】



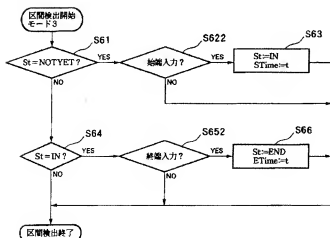
【図6】



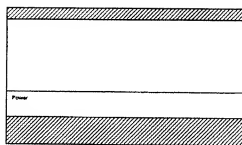
【図8】



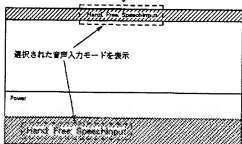
【図7】



【図9】

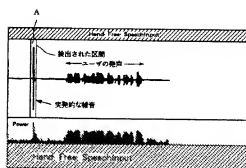


音声入力モード選択前

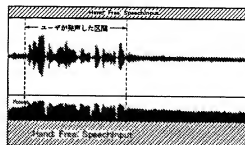


音声入力モード選択後

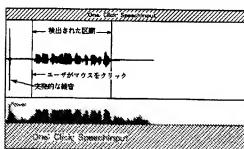
【図10】



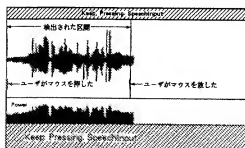
【図12】



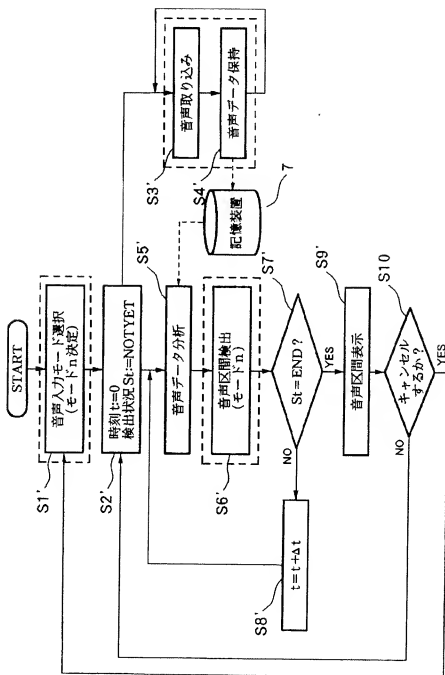
【図11】



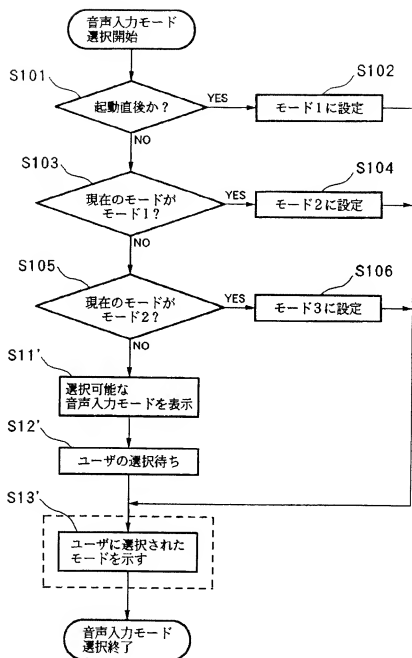
【図13】



[図14]



[図15]



フロントページの続き

(72)発明者 大洞 恭則

東京都大田区下丸子3丁目30番2号 キヤ
ノン株式会社内